

SHAMIK BASU

Los Angeles, CA | (213) 255-6219 | shamik1900@gmail.com | linkedin.com/in/shamikofficial | github.com/ShamikOfficial |
Portfolio: shamik-basu.com

SUMMARY

Data Scientist with analytics engineering and business intelligence experience turning business questions into trustworthy data models, Power BI and Tableau dashboards, SQL, and Python automation. Leading a 5-engineer team delivering a clinical document parsing model at 96% accuracy that cut manual ops by over 68%, and redesigning fraud text models for a 27% lift. Developing Machine Learning with scikit-learn, shipping AI agents on LLM APIs, and cross-collaborating with downstream business and ops.

WORK EXPERIENCE

Data Science Associate Intern | Liquidity AI Capital | Los Angeles, CA *April 2026 - May 2026*

- Built Python data processing automation (pandas, scikit-learn, Git, APIs), reaching 88% holdout and cutting p95 latency 35%.
- Designed Streamlit dashboards and documentation so non-technical partners trusted model outputs.
- Added automated reconciliation checks so mismatched numbers never reached stakeholders.

Data Science Associate Intern | KCC Capital Partners | Los Angeles, CA *January 2026 - March 2026*

- Built Python AI agents with LangChain and LLM APIs under Git, reviewing and correcting agent-written code.
- Wrote SQL with joins, CTEs, and window functions on large financial data to rank deal signals.
- Partnered with investors to turn business questions into data requirements and documentation.

Data Scientist | Bajaj Finserv Health | Pune, India *November 2023 - December 2024*

- Leading a 5-engineer team shipping production NLP (LLMs, med-spaCy, OCR) at 96% accuracy, cutting manual ops by over 68%.
- Owned Tableau and Power BI business intelligence data models with measures, refresh schedules, validation, and lineage.
- Briefed product and ops leaders on what the data can and cannot answer before launches.

Associate Data Scientist | Bajaj Finserv Health | Pune, India *July 2022 - October 2023*

- Redesigned fraud text NER models, improving fraud risk identification 27%, with clear briefings.
- Built CleverTap lead scoring with A/B tests and regression multivariate analysis (23% uplift).
- Built Salesforce lead allocation via cohort analysis and pincode geo-matching for hyperlocal coverage.

Data Engineer Intern | Bajaj Finserv Health | Pune, India *January 2022 - June 2022*

- Implemented org-level PII masking for ISO 27001 with MD5-encoded customer UUIDs for confidential data.
- Wrote SQL and Python ETL data processing on over 10M records into over 200 features with reconciliation checks.
- Partnered with modeling and ops on SQL jobs, Tableau views, Git handoffs, and testing.

EDUCATION

University of Southern California | Master's, Applied Data Science *January 2025 - December 2026*

GPA 3.7/4.0 | Relevant Coursework: Statistical Methods, Machine Learning, Foundations of AI, Business Analytics

SRM Institute of Science and Technology | Bachelor's, Computer Science *May 2018 - June 2022*

GPA 3.5/4.0 | Relevant Coursework: Operating Systems, Algorithms, Databases, Data Structures, Probability

SKILLS

Programming Languages: Python, SQL, JavaScript, TypeScript, C, C++

Machine Learning: scikit-learn, regression, time-series forecasting, statistical forecasting, A/B testing, SpaCy NER, LangChain, LLM APIs, NLP, OCR

Data: pandas, NumPy, Power BI, Tableau, business intelligence, analytics engineering, data processing, financial data, data modeling, semantic models, SQL CTEs, window functions, ETL, dataflows

Cloud Tools: Git, FastAPI, Streamlit, Docker, AWS, Azure, Salesforce, OpenAI, Claude, GitHub Copilot, coding agents

PROJECTS

Agentic Trend Orchestrator | Next.js, FastAPI, OpenAI, FAISS *May 2026 - June 2026*

GitHub: github.com/ShamikOfficial/agentic-trend-orchestrator | **Live:** trendpilot.shamik-basu.com

- Shipped coding-agent workflows with OpenAI LLM APIs and Git; post hit rate rose from 25% to 60%.
- Designed data models so agents answered live questions with tested, corrected outputs.

EcoMateAI | Streamlit, FastAPI, Gemini 2.5 Flash, Plotly *May 2025 - June 2025*

GitHub: github.com/ShamikOfficial/EcoMate-AI | **Live:** ecomateai.streamlit.app

- Built Python automation and machine learning extraction pipelines with API calls under Git.
- Delivered Plotly dashboards and documentation so non-technical users trusted ranked outputs.

HIGHLIGHTS

- Won 1st place, UCLA SAIRS Hackathon (Sustainable) with EcoMateAI.
- Led GRIDS, a USC Data Science club with over 250 members hosting workshops and hackathons.
- Mentored incoming graduate students on USC culture and campus resources.
- Contributed 6 merged PRs and 18 open PRs across Hugging Face, NetworkX, Hydra, Tornado, NumPy, and python-dotenv.